Figure 1.

```
Sequence type explicitly set to Protein
Sequence format is Pearson
Sequence 1: new_S100_cytokine        104 aa
Sequence 2: G491246                  110 aa
Sequence 3: W27152                    98 aa
Start of Pairwise alignments
Aligning...
Sequences (1:2) Aligned. Score:  23
Sequences (1:3) Aligned. Score:  34
Sequences (2:3) Aligned. Score:  29
Start of Multiple Alignment
There are 2 groups
Aligning...
Group 1:                    Delayed
Group 2:                    Delayed
Sequence:1      Score:0
Sequence:3      Score:839
Sequence:2      Score:724
Alignment Score 444
CLUSTAL-Alignment file created   [/data4/genetools/lrastelli4630clustalw]
```

Multiple Alignment:

```
new_S100_cytokine   MGQCRSANAEDAQEFSDVERAIETLIKNFHQYSVEGG-KETLTPSELRDLVTQQLPHLMP
W27152              ---------MAAEPLTELEESIETVVTFFTFARQEGRKDSLSVNEFKELVTQQLPHLLK
G491246             --------------MSQLERNIETIINTFHQYSVKLGHPDTLNQGEFKELVRKDLQNFLK

new_S100_cytokine   SNCG----LEEKIANLGSCNDSKLEFRSFWELIGEAAKSVKLERP----VRGH---(SEQ ID NO:6)
W27152              DVGS----LDEKMKSLDVNQDSELKFNEYWRLIGELAKEIRKKKD----LKIRKK-(SEQ ID NO:10)
G491246             KENKNEKVIEHIMEDLDTNADKQLSFEEFIMLMARLTWASHEKMHEGDEGPGHHHKPGLG

new_S100_cytokine   - - - -
W27152              - - - -
G491246             EGTP (SEQ ID NO:11)
```

Figure 2.

Multiple Alignment:

```
new_S100_cytokine       --------------DNRTLTKGPDTVS-TMGQCRSANAEDAQEFSDVERAIETLIKNFHQYS
7971c.7___r0s0-212.2__EXT  SISSCGAGYRTDDKTQLTEGRTSVPGTMGQCRSANAEDAQEFSDVERAIETLIKNFHKYS

new_S100_cytokine       VEGGKETLTPSELRDLVTQQLPHLMPSNCGLEEKIANLGSCNDSKLEFRSFWELIGEAAK
7971c.7___r0s0-212.2__EXT  VAGKKETLTPAELRDLVTQQLPHLMPSNCGLEEKIANLGNCNDSKLEFGSFWELIGEAAK

new_S100_cytokine       SVKLERPVRGH    (SEQ ID NO:6)
7971c.7___r0s0-212.2__EXT  SVKMERPVTRS    (SEQ ID NO:3)
```

Figure 3.

```
hmmpfam - search a single seq against HMM database


HMMER is freely distributed under the GNU General Public License (GPL).
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
HMM file:                 pfamHMMs
Sequence file:            /data4/genetools/1rastelli4423Aa315020ProteinFasta.txt
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Query:  AA315020


Scores for sequence family classification (score includes all domains):
Model      Description                                     Score    E-value  N
--------   -----------                                     -----    -------  ---

S_100      S-100/ICaBP type calcium binding domain         40.9     2.9e-08  1


Parsed for domains:
Model      Domain  seq-f  seq-t     hmm-f  hmm-t      score  E-value
--------   ------- -----  -----     ----- -----      -----  -------

S_100      1/1       32     74 ..       1     44 []    40.9  2.9e-08


Alignments of top-scoring domains:
S_100: domain 1 of 1, from 32 to 74: score 40.9, E = 2.9e-08
                   *->LEkaietiInvFhqYSgreGdkdtLsKkELKeLlekELpnfLkn<-*
                      E+aiet+I+ FhqYS  eG k tL+   EL+ L++++Lp+ +
       AA315020    32    VERAIETLIKNFHQYS-VEGGKETLTPSELRDLVTQQLPHLMPS      74(SEQ ID NO:33)


//
```

Figure 4A.

BLOCKS Protein Domain Analysis of new_S100_cytokine

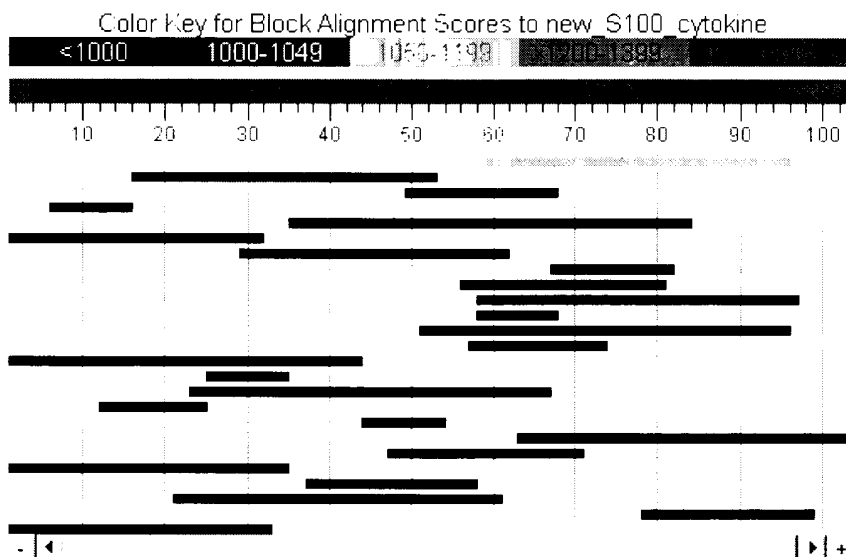Probe Size: 104 Amino Acids

Probe File: lrastelliblocks.seq

Target File (s) : blocks.dat

Records Searched: 4034

Scores Done: 4034

Alignments Done: 535470

Mouse-over to show defline and scores. Click to show alignments

Color Key for Block Alignment Scores to new_S100_cytokine

| <1000 | 1000-1049 | 1050-1199 | | |

AC#	Description	Strength	Score	HT	AA#	
BL00?0?B	S-100/ICaBP type calcium binding protein.	1236	1057	0	59 sNCGLEEKiaNLgSCrDSKLeFRSFWeLIGEAAKsVk	(SEQ ID NO:12)
BL00?0? A	S-100/ICaBP type calcium binding protein.	1245	1028	0	16 DvERAIETLIKnFHQYSVEgGKeTltpsElrdLvtQQ	(SEQ ID NO:13)
BL00074A	Bacterial type II secretion system protein F	1456	1031	0	43 VTQQLphLMpSncgLEEki	(SEQ ID NO:14)
BL0037.B	Ubiquitin carboxyl-terminal hydrolases family	1227	391	0	6 AnaEDAQEFs	(SEQ ID NO:15)
BL00?5?6	Bacterial chemotaxis sensory transducers prot	1258	380	0	35 GGkEtlTpselRD1VTQQ1phLmpsncgleEkiANlgScndSkleFRsf	(SEQ ID NO:16)
BL00?3?I	Phosphoenolpyruvate carboxykinase (ATP) prote	1412	389	0	mgQtRSAnAedAQefS dVERaIEtlIKNFhqY	(SEQ ID NO:17)
BL00?04A	Prokaryotic-type carbonic anhydrases proteins	1529	3x?	0	29 hqYSvegGKetltPselrdLvTQQ1PHlMpsnC	(SEQ ID NO:18)
BL01017E	Ergosterol biosynthesis ERG4/ERG24 family pro	1499	9??	0	67 ianlgSCNdsKlEFR	(SEQ ID NO:19)
BL00?19E	Lysosome-associated membrane glycoproteins du	1622	976	0	56 LMp3NcgLeEkiAnLGScndsklEf	(SEQ ID NO:20)
BL0043?0	Phosphofructokinase proteins.	1581	976	0	56 p3nCGLeekIaNlgSCrDSkleFrSfwElIgEAaKSVkl	(SEQ ID NO:21)
BL5000?	PH domain proteins profile	990	976	0	56 PSNCGLEEKI	(SEQ ID NO:22)
BL00453	Myotoxins proteins.	2172	974	0	51 qqlpHlmpsnCgleEKIanlgScndsXlefRsfWeligeaaksvk	(SEQ ID NO:23)
BL5000?0	Phosphatidylinositol-specific phospholipase X	142?	974	0	57 mP3nCgLeEKIanlbsC	(SEQ ID NO:24)
BL01?0?E	Glypicans proteins.	1508	97?	0	-1 mggCrSAMAEDaQeFsDVeRaIeTlIKnfhQySveggketLTpSE	(SEQ ID NO:25)
BL00?24E	Membrane attack complex components / perforin	1127	97?	0	23 TLIknfhqySveggkeTlTPSElrdlvtQQlphLMPsNcGLeeX	(SEQ ID NO:26)
BL0112?E	Urease nickel ligands proteins.	1692	970	0	12 qefSDvERAIETL	(SEQ ID NO:27)
BL001?0E	Phosphoglycerate mutase family phosphohistidi	1298	968	0	44 ELRDLvTqQL	(SEQ ID NO:28)
BL00?29A	Ribosomal protein L29 proteins.	1092	967	0	63 LeekiaNLgScnDsKlEfrsFweliIgEaaKsVklERpvRgh	(SEQ ID NO:29)
BL00??;E	2'-5'-oligoadenylate synthetases proteins.	1876	967	0	47 dLVTqqLPhLMpsntGlEekIAML	(SEQ ID NO:30)
BL0100?C	Formate and nitrite transporters proteins.	1552	964	0	1 gQCRsANaedaqEFsDVerAIETlIknfHqYSvE	(SEQ ID NO:31)
BL00?61A	Glycoprotein hormones beta chain proteins.	1528	963	0	33 keTLTpseLrd1VtQQ1phLM	(SEQ ID NO:32)
BL00?66 A	Vinculin family talin-binding region proteins	1567	963	0	21 IeTlIkNfhqysveGGkETLtpselrdlvTQQlphLMpSN	(SEQ ID NO:33)
BL0074 A	Beta-lactamases class B proteins.	1580	960	0	78 LEFRSFwELIGEAAKSVKLER	(SEQ ID NO:34)
BL01031B	Heat shock hsp20 proteins family profile.	1200	960	0	-2 MggcrsAnAedAqEfSdverAiEtliKNfhqySvE	(SEQ ID NO:35)
BL0106?0	Hydroxymethylglutaryl-coenzyme A lyase protei	1908	960	0		(SEQ ID NO:36)

Figure 4B.

```
                 10        20        30        40        50        60
        ....|....|....|....|....|....|....|....|....|....|....|....|
Table3    MGQCRSANAEDAQEFSDVERAIETLIKNFHKYSVAGKKETLTPAELRDLVTQQLPHLMPS
AA007220  MGQCRSANAEDAQEFSDVERAIETLIKNFHQYSVEGGKETLTPSELRDLVTQQLPHLMPS
Consensus MGQCRSANAEDAQEFSDVERAIETLIKNFH YSV G KETLTP ELRDLVTQQLPHLMPS

                 70        80        90       100
        ....|....|....|....|....|....|....|....|....|....|..
Table3    NCGLEEKIANLGNCNDSKLEFGSFWELIGEAAKSVKMERPV      (SEQ ID NO:3)
AA007220  NCGLEEKIANLGSCNDSKLEFRSFWELIGEAAKSVKLERPV      (SEQ ID NO:39)
Consensus NCGLEEKIANLG CNDSKLEF SFWELIGEAAKSVK ERPV      (SEQ ID NO:40)
```

Figure 4C.

```
                 10        20        30        40        50        60
        ....|....|....|....|....|....|....|....|....|....|....|....|
Table 6   MGQCRSANAEDAQEFSDVERAIETLIKNFHQYSVEGGKETLTPSELRDLVTQQLPHLMPS
AA007220  MGQCRSANAEDAQEFSDVERAIETLIKNFHQYSVEGGKETLTPSELRDLVTQQLPHLMPS
Consensus MGQCRSANAEDAQEFSDVERAIETLIKNFHQYSVEGGKETLTPSELRDLVTQQLPHLMPS

                 70        80        90       100
        ....|....|....|....|....|....|....|....|....|....
Table 6   NCGLEEKIANLGSCNDSKLEFRSFWELIGEAAKSVKLERPVRGH    (SEQ ID NO:6)
AA007220  NCGLEEKIANLGSCNDSKLEFRSFWELIGEAAKSVKLERPVRGH    (SEQ ID NO:39)
Consensus NCGLEEKIANLGSCNDSKLEFRSFWELIGEAAKSVKLERPVRGH    (SEQ ID NO:40)
```

Figure 4D.

```
                    10        20        30        40
           ....|....|....|....|....|....|....|....|....|.
Table 3        ERAIETLIKNFHKYSV-AGKKETLTPAELRDLVTQQLPHLM   (SEQ ID NO:3)
gi|4139958|pdb|1MHO|  EKAVVALIDVFHQYSGREGDKHKLKKSELKELINNELSHFL   (SEQ ID NO:41)
PROTEIN MRP-126  EKAIDVIIDVFHQYSRREGDKDTLTRKELKLLIEKQLANYL   (SEQ ID NO:42)
ICTACALCIN     QKGMAILISTFHKYSGKEGDKCTLTKGELKDLLTKELGGAF   (SEQ ID NO:43)
CALGRANULIN B  ESSIETIINIFHQYSVRLGHYDTLIQKEFKQLVQKELPNFL   (SEQ ID NO:44)
Consensus          I  FHYS  G    L  E    L        L     (SEQ ID NO:45)
```

Figure 4E.

```
                    10        20        30        40
           ....|....|....|....|....|....|....|....|....|.
Table 6        ERAIETLIKNFHQYSV-EGGKETLTPSELRDLVTQQLPHLM   (SEQ ID NO:6)
gi|4139958|pdb|1MHO|  EKAVVALIDVFHQYSGREGDKHKLKKSELKELINNELSHFL   (SEQ ID NO:41)
PROTEIN MRP-126  EKAIDVIIDVFHQYSRREGDKDTLTRKELKLLIEKQLANYL   (SEQ ID NO:42)
CALGRANULIN B  ESSIETIINIFHQYSVRLGHYDTLIQKEFKQLVQKELPNFL   (SEQ ID NO:44)
CALGRANULIN B  ERSITTIILDTFHQYSRKEGHPDTLSKKEFRQMVEAQLATFM  (SEQ ID NO:46)
Consensus      E    I   FHQYS  G    L  E        L        (SEQ ID NO:47)
```
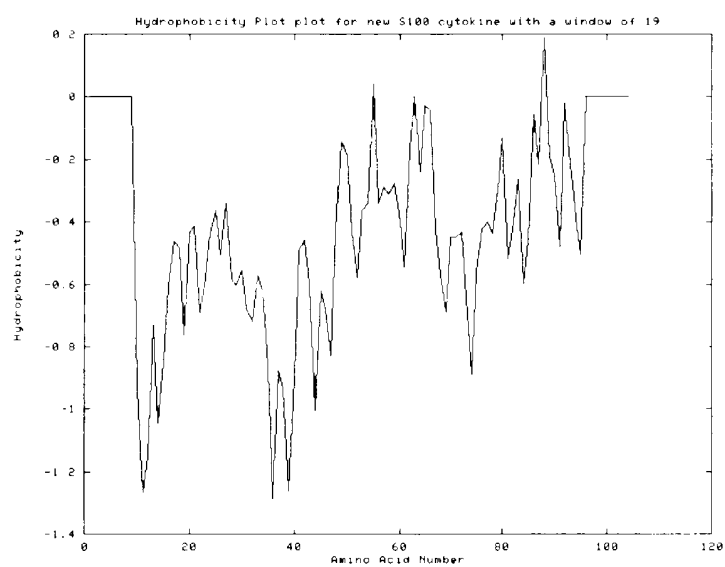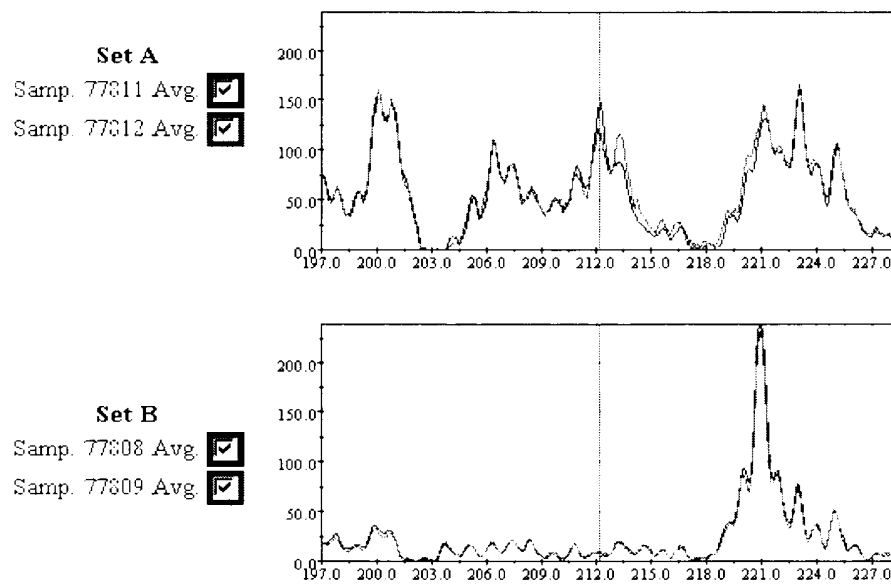
Figure 5

Hydrophobicity Plot plot for new S100 cytokine with a window of 19

Figure 6

Set A

Samp. 77811 Avg. ☑
Samp. 77812 Avg. ☑

Set B

Samp. 77808 Avg. ☑
Samp. 77809 Avg. ☑

Figure 7

```
****************** Contig 1 ******************
           .    :    .    :    .    :    .    :    .    :
65677221+  GAATTCCAGAGGGAGTTCTCAGTGCCCCCGGACAGGCCTCTCCAGCTTCACACTCTTGGC
AA315020-                    TGCCCCCGGACAGTCCTCTCNAGCTTCACACTCTTGGC
consensus  GAATTCCAGAGGGAGTTCTCAGTGCCCCCGGACAGGCCTCTCCAGCTTCACACTCTTGGC

           .    :    .    :    .    :    .    :    .    :
65677221+  CGCTTCTCCAATCAGCTCCCAGAAACTCCTGAACTCCAGTTTAGAGTCATTGCAGCTGCC
AA315020-  CGCTTCTCCAATCAGCTCCCAGAAACTCCTGAACTCCAGTTTAGAGTCATTGCAGCTGCC
consensus  CGCTTCTCCAATCAGCTCCCAGAAACTCCTGAACTCCAGTTTAGAGTCATTGCAGCTGCC

           .    :    .    :    .    :    .    :    .    :
65677221+  CAGGTTGGCAATTTTCTCTTCCAGGCCACAGTTGCTCGGCATGAGATGGGGCAGCTGCTG
AA315020-  CAGGTTGGCAATTTTCTCTTCCAGGCCANAGTTGCTCGGCATGAGATGGGGCAGCTGCTG
consensus  CAGGTTGGCAATTTTCTCTTCCAGGCCACAGTTGCTCGGCATGAGATGGGGCAGCTGCTG

           .    :    .    :    .    :    .    :    .    :
65677221+  GGTGACCAGGTCCCGTAGCTCAGAAGGGGTCAGCGTCTCCTTCCCACCCTCCACGGAGTA
AA315020-  GGTGACCAGGTCCCGTAGCTCAGAAGGGGTCAGCGTCTCCTTCCCACCCTCCACGGAGTA
consensus  GGTGACCAGGTCCCGTAGCTCAGAAGGGGTCAGCGTCTCCTTCCCACCCTCCACGGAGTA

           .    :    .    :    .    :    .    :    .    :
65677221+  CTGGTGAAAGTTCTTGATGAGGGTCTCAATGGCCCTCTCCACATCACTGAATTC (SEQ ID NO:37)
AA315020-  CTGGTGAAAGTTCTTGATGAGGGTCTCAATGGCCCTCTCCACATCACTGAATTCCTGAGC
consensus  CTGGTGAAAGTTCTTGATGAGGGTCTCAATGGCCCTCTCCACATCACTGAATTCCTGAGC

           .    :    .    :    .    :    .    :    .    :
AA315020-  ATCCTCTGCGTTGGCTGACCGACACTGTCCCATGGTGCTCACTGTGTCTGGTCCTTTGGT
consensus  ATCCTCTGCGTTGGCTGACCGACACTGTCCCATGGTGCTCACTGTGTCTGGTCCTTTGGT

           .    :    .    :    .    :    .    :    .    :
AA315020-  GAGAGTTCTGTTGTCCTAT    (SEQ ID NO:4)
consensus  GAGAGTTCTGTTGTCCTAT    (SEQ ID NO:5)
```